2025 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON)
29-30 November 2025, Eastern University, Dhaka, Bangladesh

2025 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON) | 979-8-3315-6105-5/25/$31.00 ©2025 IEEE | DOI: 10.1109/BECITHCON69222.2025.11503998

SkinViT-EfficientX: A Hybrid Vision Transformer Model with Token Pruning and Explainable AI for Multiclass Skin Cancer Diagnosis

Mostafizur Rahman Shakil[1], Mahfuzur Rahman[2], Erin Jahan Meem[3], Md Imranul Hoque Bhuiyan[4], Sanjida Akter[5], Arafath Bin Mohiuddin[6], Shafiur Rahman[7], Istiak Kabir[8]
[1]*Department of Engineering Management, Westcliff University, CA 92614, USA*
[2]*School of Computing and Digital Media, London Metropolitan University, London N7 8DB, UK*
[3]*Department of Computer Science, Pacific States University, CA 90010, USA*
[4]*Department of Business Analytics, International American University, CA 90010, USA*
[5]*Department of Mathematics and Natural Sciences, BRAC University, Dhaka, Bangladesh*
[6]*Department of Information Technology, Westcliff University, CA 92614, USA*
[7]*Dept. of Software Engineering, Daffodil International University, Dhaka 1216, Bangladesh*
[8]*Department of Computer Science, Westcliff University, CA 92614, USA*
Email: [1]mostafizurrahmanshakil@gmail.com, [2]mahfuzurrahman63036@gmail.com, [3]erinjahanmim@gmail.com, [4]imranrafi1111@gmail.com,[5]sanjidairen261@gmail.com, [6]arafath2307@gmail.com, [7]shafiur016@gmail.com, [8]istiak2221@gmail.com

***Abstract*— Skin cancer is a common and serious health issue, making early diagnosis crucial for better outcomes. Traditional manual dermoscopy can be slow and inconsistent, demonstrating a need for automated diagnostic tools. This study introduces SkinViT-EfficientX, a hybrid deep learning model specifically designed for classifying skin lesions. It utilizes an EfficientNetV2-S encoder and a lightweight Vision Transformer connected by a residual cross-attention mechanism for effective local-global feature extraction. To enhance performance, a confidence-guided token pruning strategy is employed, and Grad-CAM is used for class-specific visual explanations. The model underwent thorough preprocessing and augmentation on two benchmark datasets: HAM10000 and the combined ISIC 2019 + DermNet dataset. SkinViT-EfficientX achieved a 97.36% F1-Score, 95.64% MCC, and 97.93% Specificity on HAM10000, while scoring 98.42% F1-Score, 96.51% MCC, and 98.86% Specificity on the combined dataset. It outperformed top models like MaxViT, Swin V2-T, DeiT III-S, and MobileViT V2-S in all metrics. The model's robustness and stability for rare lesion classes were validated through confusion matrix and learning curve analyses. Further, it is integrated into a web application for dermoscopic image uploads, class predictions, and heatmap visualizations. SkinViT-EfficientX provides an efficient, accurate, and interpretable AI-driven solution for skin cancer screening.**

Keywords—Skin cancer, vision transformer, explainable AI, medical imaging, diagnostic tools

I. INTRODUCTION

Skin cancer, including melanoma and non-melanoma types, is a major global health issue due to its rising incidence and impact on healthcare. The World Health Organization (WHO) reports over 3 million non-melanoma skin cancer cases and around 132,000 melanoma cases diagnosed annually [1]. In high-risk areas like North America and Oceania, the lifetime risk of developing skin cancer is over 20%, with melanoma causing most deaths. Early detection significantly improves outcomes; for example, the 5-year survival rate for early diagnosed melanoma patients can exceed 99% [2]. However, late-stage diagnoses remain common, especially among underserved communities. Dermoscopic examination is the clinical gold standard for skin lesion assessment, but it is inherently subjective and requires extensive expertise. This subjectivity limits scalability and consistency across different practitioners and institutions.

Recent advances in artificial intelligence (AI) and deep learning have enabled automated classification of skin lesions using dermoscopic images. Convolutional Neural Networks (CNNs) have shown promise in capturing texture, color, and shape features directly from image data [3]. However, they often struggle with overfitting, limited generalizability across devices, and high computational demands. Many models are also restricted to binary classification and lack interpretability, an essential factor for clinical acceptance. Vision Transformers (ViTs) have emerged as strong alternatives, leveraging self-attention mechanisms to capture global dependencies. While ViTs achieve high accuracy, they require substantial computational resources and often lack spatial inductive bias, which is important for recognizing local lesion features. Additionally, they typically lack built-in explainability and are not optimized for real-time or mobile deployment, limiting their practical applicability.

To address these challenges, this study sets out four key objectives: (1) to develop a lightweight hybrid model that combines CNN-based local feature extraction with ViT-based global reasoning for multi-class skin cancer classification; (2) to reduce computational complexity through confidence-guided token pruning; (3) to enhance transparency using integrated Grad-CAM explainability; and

(4) to deploy the model in a real-time, web-based application for accessible and scalable clinical use.

To achieve these objectives, we present SkinViT-EfficientX, a hybrid model that combines an EfficientNetV2-S convolutional stem with a lightweight transformer encoder and residual cross-attention fusion (Fig. 1). It uses multi-scale tokenization, convolutional positional encoding, and a pruning mechanism to eliminate low-importance tokens, reducing computational load while maintaining accuracy. Integrated Grad-CAM produces class-specific saliency maps from the fusion layer for better model transparency. Testing on multiple datasets shows that SkinViT-EfficientX outperforms baseline ViTs in various metrics. The core contributions of this work are:

- Proposed a hybrid model that combines EfficientNetV2-S with a lightweight Vision Transformer using residual cross-attention fusion and confidence-guided token pruning.
- Utilized two large-scale dermoscopic datasets, HAM10000 and a Combined ISIC 2019 + DermNet dataset, ensuring class balance and improved model generalization through stratified splitting and advanced augmentation.
- SkinViT-EfficientX outperforms existing ViT and CNNs across various evaluation metrics, showing strong and consistent performance on both datasets.
- Developed a real-time web-based interface that integrates Grad-CAM for visualizing lesion-focused areas, enhancing clinical transparency.

The paper is organized as follows: Section 2 reviews related work, Section 3 outlines the methodology, Section 4 presents results, Section 5 discusses key findings, and Section 6 concludes with future directions.

II. Related Works

AI applications are diverse, spanning legal analysis, cybersecurity, spatial positioning, network science, deliberation systems, and blockchain exchanges [4], [5], [6]. In agriculture and transportation, AI-driven navigation and segmentation have shown great effectiveness [7], [8], [9]. In industrial settings, reinforcement learning and multi-modal learning enhance automation and predictive analytics [10], [11], [12], [13]. In healthcare, advanced models improve the classification of neurological and cancer diseases [14], [15]. Additionally, AI techniques are applied in IoT security and heat transfer systems [16], [17], [18], impacting engineering and environmental sciences.

Bechelli et al. [19] found that the VGG16 deep learning model outperformed traditional machine learning methods in skin tumor classification, achieving 88% accuracy. However, it has high computational demands and tends to overfit small datasets, highlighting the need for models that generalize better. To tackle this, Tahir et al. [20] introduced DSCCNet, which achieved an impressive AUC of 99.43%, 94.17% accuracy, and 93.93% F1-score across three benchmark datasets. Still, its real-world clinical validation and generalizability are uncertain due to potential dataset-specific overfitting. Jain et al. [21] evaluated six transfer learning networks on the HAM10000 dataset, finding that XceptionNet led with 90.48% accuracy, but did not explore fine-tuning methods to boost diagnostic accuracy. Farea et al. [22] proposed a hybrid learning framework using multiple public datasets for training CNNs, achieving 93.04% accuracy and 93.12% F1-score while employing the artificial bee colony algorithm, although this adds computational complexity. Moreover, its reliance on pre-trained models limits adaptability to new datasets. Panthakkan et al. [23] developed a hybrid X-R50 model combining Xception and ResNet50 for classifying melanoma, BCC, and keratoacanthoma, achieving 97.8% accuracy on 10,500 images but introducing computational overhead.

Recent studies have incorporated explainable AI (XAI) to improve interpretability. Grignaffini et al. [24] used a CNN with LBP feature injection and various XAI techniques, reaching 98.41% accuracy, though dependence on handcrafted features limits clinical utility. Abdulredah et al. [25] introduced SWNet, achieving 99.86% accuracy and 99.95% F1-score but requiring extensive preprocessing and having issues with dataset homogeneity. Halder et al. [26] utilized a fuzzy ensemble of models, reaching 94.94% accuracy on Document Number 1 and 78.25% on DermaMNIST, facing challenges with computational efficiency. Similarly, Cino et al. [27] achieved 97.58% balanced accuracy with EfficientNet-B6 by using test-time augmentation and Grad-CAM but lacked validation across different devices.

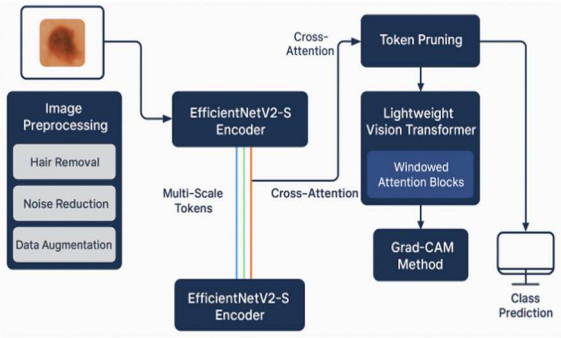

Fig. 1. Proposed architecture of this study.

While recent studies show high accuracy but face challenges such as high computational costs, overfitting, and insufficient clinical validation. Many models have difficulty generalizing across various imaging conditions and often require extensive preprocessing. While XAI tools like Grad-CAM enhance interpretability, they usually focus on non-clinical features and lack multi-center or cross-device validation. Furthermore, most methods either focus on binary classification or are too complex for deployment.

III. Methodology

A. Data Description

We employed two publicly available dermoscopic image datasets: HAM10000 [28] and a combined dataset integrating ISIC 2019 [29] and DermNet [30] dataset. HAM10000 includes 10,015 images annotated across seven lesion categories: Melanocytic Nevi (NV, 6,705), Melanoma (MEL, 1,113), Benign Keratosis-like Lesions (BKL, 1,099), Basal Cell Carcinoma (BCC, 514), Actinic Keratoses (AKIEC, 327), Vascular Lesions (VASC, 142), and Dermatofibroma (DF, 115), offering clinical diversity in lesion type, location, and demographics. The combined dataset comprises 30,990 images, merging 25,331 samples from ISIC 2019 and 5,659 from DermNet, spanning eight classes: NV (12,875), MEL (6,102), BK (4,338), BCC (3,323), AK (2,304), Squamous

Cell Carcinoma (SCC, 628), VASC (856), and DF (564). Both datasets were partitioned using stratified sampling into 80% training, 5% validation, and 15% testing splits. Sample image is displayed in Fig. 2.

B. Data Preprocessing

All dermoscopic images were resized to 224×224 pixels using bilinear interpolation for model compatibility. Pixel intensities were normalized to the [0,1] range to stabilize training and improve convergence in multi-class classification. The Dull Razor algorithm was applied to remove hair artifacts while preserving lesion structures through grayscale morphological operations and adaptive median filtering. Anisotropic diffusion filtering reduced background noise and preserved lesion edges by adjusting diffusion coefficients based on local gradients, crucial for differentiating skin cancer types. To enhance data diversity and model generalization, the training set was augmented with techniques such as random rotation (±25°), flipping (50% chance), brightness (±20%), contrast (±15%), zoom scaling (90–110%), shear (±20%), cropping (10–20%), and elastic stretching (3–5 pixel grid deformation). This increased the training set to approximately 64,000 images, with 8,000 samples per class across eight categories. The HAM10000 dataset was also augmented to 14,000 images, ensuring a balanced representation of 2,000 samples per class, including rare categories like SCC, DF, and VASC.

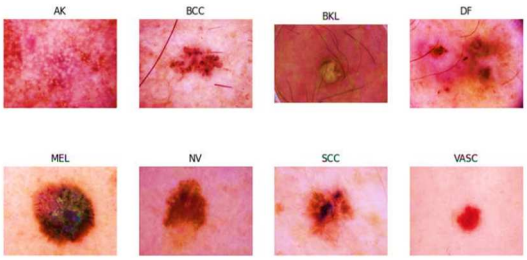

Fig. 2. Sample image from datasets.

C. Base Models

We selected four recent ViT architectures as baseline models. These models were chosen for their complementary strengths in capturing local and global dependencies, computational efficiency, and adaptability to medical imaging tasks.

MaxViT-T integrates both convolutional and attention-based operations in a hierarchical structure. It applies depthwise convolution followed by factorized attention blocks to capture local texture and global structure in dermoscopic images. The depthwise convolution operation used in MaxViT is defined as Eq. 1, where (X) is the input feature map, (W) is, the filter, and (c) indexes the channel. MaxViT is effective for identifying complex lesion borders and regional irregularities [31].

$$Y_{i,j,c} = \sum_{m=-k}^{k} \sum_{n=-k}^{k} W_{m,n,c} \cdot X_{i+m,j+n,c} \quad (1)$$

Swin Transformer V2-T adopts a shifted window self-attention mechanism that enhances spatial locality and reduces computation through partitioned attention. It maintains a hierarchical representation through patch merging, which aggregates nearby pixel patches to lower spatial resolution and increase channel dimensionality [32]. The patch merging operation is expressed as Eq. 2. This structure enables multi-scale feature extraction, making Swin V2 suitable for detecting lesions with variable sizes, shapes, and distributions. However, the fixed window size may limit its ability to capture long-range dependencies across distant regions.

$$Z_{i,j} = \text{Linear}\left(\left[X_{2i,2j}, X_{2i,2j+1}, X_{2i+1,2j}, X_{2i+1,2j+1}\right]\right) \quad (2)$$

DeiT III-S is optimized for low-data regimes, which is common in medical imaging. It enhances the standard ViT by introducing a distillation token that interacts with transformer layers to absorb teacher supervision during training. The forward update of this distillation token is defined as Eq. 3, where (LN(·)) denotes layer normalization and (MLP(·)) represents a feedforward network. This design allows DeiT III to generalize better from small datasets. Nevertheless, its lack of locality-preserving operations may reduce sensitivity to fine-grained lesion patterns unless supplemented by extensive regularization or augmentation.

$$d_t^{(l+1)} = d_t^{(l)} + \text{MLP}\left(\text{LN}\left(d_t^{(l)}\right)\right) \quad (3)$$

MobileViT V2-S features a hybrid design, combining the efficiency of convolutions with the global reasoning capabilities of transformers. It first projects feature maps into sequences using reshaping, apply transformer operations, and then reverts them to spatial format [33]. The overall process is represented by Eq. 4. Due to its low parameter count and fast inference time, it is ideal for deployment in web-based clinical applications. However, its reduced representational capacity may lead to underfitting on complex or ambiguous lesion types, particularly in cases with overlapping color and shape distributions.

$$Y = \text{Conv}_{1\times1}\left(\text{Reshape}^{-1}\left(\text{Transformer}\left(\text{Reshape}(X)\right)\right)\right) \quad (4)$$

D. Proposed SkinViT-EfficientX

ViT models like MaxViT-T and Swin V2-T perform well but require high computational power and lack real-time interpretability for clinical use. In contrast, lightweight models like MobileViT V2-S offer quicker inference but may struggle with complex dermatological features. To address these challenges, we present SkinViT-EfficientX, a hybrid deep learning model that incorporates convolutional inductive bias, global attention, token pruning, and explainable AI for effective skin cancer classification in a compact format. SkinViT-EfficientX begins in Fig. 3 with an EfficientNetV2-S stem, which acts as a lightweight convolutional encoder to extract high-resolution local features critical for maintaining lesion boundaries and texture. The output feature maps go through a convolutional positional encoding block to preserve spatial structure and are reshaped into a hierarchical token sequence using a multi-scale tokenizer. A lightweight ViT encoder, inspired by the Swin Transformer, processes these tokens to efficiently capture both low-level details and high-level patterns. To combine the outputs from the CNN and transformer, we implement a residual cross-attention fusion mechanism. This mechanism uses the EfficientNetV2 output and the token embeddings from the transformer layer, controlled by a learnable parameter that balances the contribution of local CNN features. This design preserves texture fidelity while enabling global context modeling.

To further reduce computational burden, SkinViT-EfficientX incorporates a confidence-guided token pruning

strategy. For each token ($t_i \in T$), its importance is estimated based on attention confidence. Tokens with a confidence score below a defined threshold (τ) are removed prior to classification. The pruning operation is formulated as follows Eq. 5.

$$T' = \{t_i \in T \mid \text{Conf}(t_i) > \tau\} \quad (5)$$

This mechanism enables the model to discard irrelevant tokens, thereby reducing floating point operations (FLOPs) and enhancing inference efficiency without significant loss in accuracy. The pruned token set is then passed through a global average pooling layer and a fully connected classification head to produce probability scores for seven skin cancer classes. To enhance interpretability, Grad-CAM is integrated into the fusion layer, allowing for the generation of attention heatmaps that highlight critical lesion regions used for prediction. This feature increases the model's trustworthiness in clinical decision-making scenarios.

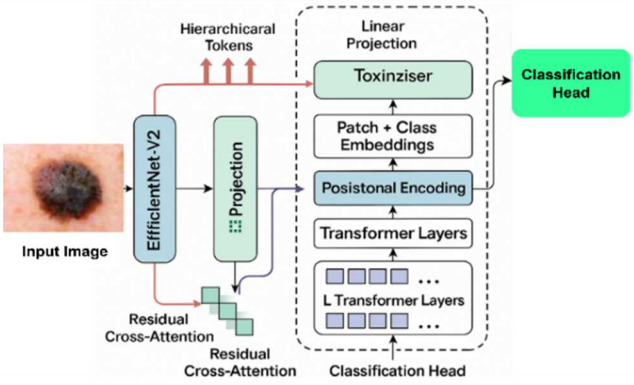

Fig. 3. Proposed SkinViT-EfficientX.

SkinViT-EfficientX offers several advantages: it captures both local and global patterns, reduces computational costs through token pruning, improves model explainability via built-in Grad-CAM, and maintains high performance on limited hardware. Compared to baseline models, it achieves a superior balance between accuracy and transparency.

E. Evaluation

We used four key metrics for evaluation: F1-Score, Precision-Recall Area Under the Curve (PR-AUC), Matthews Correlation Coefficient (MCC), and Specificity. These metrics were chosen to ensure a robust assessment of models dealing with imbalanced data while highlighting both predictive quality and clinical significance. The F1-Score balances precision and recall, which is crucial in scenarios where false negatives and positives can have serious consequences. PR-AUC measures how well a model maintains precision and recall at different thresholds, reflecting its sensitivity and robustness. MCC offers a comprehensive view of classification quality by considering all categories of the confusion matrix, making it useful for multi-class setups. Specificity, focusing on true negatives, is vital in medical contexts to prevent unnecessary treatment or anxiety for patients with non-cancerous conditions. In addition, we analyzed learning curves to observe model convergence and training dynamics. We also reviewed confusion matrices to evaluate per-class performance, identifying skin cancer categories that are frequently misclassified.

IV. Result Analysis

A. Comparing the Performance of Models

Table I compares five Vision Transformer models using metrics from HAM10000 on a Combined Dataset. SkinViT-EfficientX outperformed all others, achieving an F1 score of 97.36 and an MCC of 95.64, indicating strong accuracy and robustness. MobileViT V2-S followed closely but had lower MCC and Specificity, suggesting it's less stable with class imbalances. DeiT III-S excelled in PR-AUC at 97.29 but had a lower MCC of 90.42, indicating decision boundary inconsistencies. MaxViT and Swin Transformer V2-T showed moderate but balanced performance.

TABLE I. Performance of The Different Classifiers

HAM10000				
Model	F1	PR-AUC	MCC	Specificity
MaxViT	93.67	95.08	91.11	93.73
Swin V2-T	94.38	95.23	92.06	94.57
DeiT III-S	96.81	97.29	90.42	92.76
MobileViT V2-S	95.13	96.74	93.61	95.41
SkinViT-EfficientX	97.36	98.64	95.64	97.93
Combined Dataset				
SwinV2-T	94.18	96.14	92.21	94.63
MaxViT	95.82	96.13	91.96	95.62
MobileViT V2-S	93.51	93.91	91.05	93.28
DeiT III-S	97.26	98.09	94.41	97.48
SkinViT-EfficientX	98.42	99.02	96.51	98.86

In the Combined Dataset analysis, SkinViT-EfficientX again led with a PR-AUC of 99.02 and Specificity of 98.86, demonstrating superior generalization. DeiT III-S improved across all metrics, showing better adaptability. MaxViT saw slight gains in F1 and PR-AUC but still had a low MCC. MobileViT V2-S continued to be the weakest in Specificity and MCC, confirming limited robustness. Swin Transformer V2-T had stable but unimpressive results in both datasets.

B. Performance Validation

Fig. 4 presents the confusion matrices for the SkinViT-EfficientX model on the 15% test splits of HAM10000 and a Combined Dataset (ISIC 2019 + DermNet). In the HAM10000 matrix, the model shows strong performance, particularly in the NV class with 855 correct classifications and 141 for MEL, though there are some misclassifications with similar lesions like BKL, BCC, and AKIEC. Rare classes (VASC, DF) exhibit slight confusion, but overall specificity is high at 97.93%. In the Combined Dataset, performance remains consistent across eight classes, achieving high true positives for NV (1641) and MEL (777). Underrepresented classes like SCC and DF are classified accurately with minimal errors, mostly occurring between visually similar categories (e.g., BK vs. BCC). The model maintains a strong diagonal structure with a specificity of 98.42%.

Fig. 5 shows that the model performs well on both datasets. The Combined dataset exhibits better stability and generalization due to its larger size and diversity. There are no signs of underfitting or severe overfitting. In the HAM10000 dataset, training loss decreases steadily, with validation loss also declining and showing minor fluctuations after epoch 10, indicating occasional overfitting but overall stable convergence. Both training and validation accuracy remain above 95%, demonstrating strong generalization. The Combined dataset has smoother loss curves with minimal

divergence, and validation accuracy closely follows training accuracy, highlighting effective learning and excellent generalization.

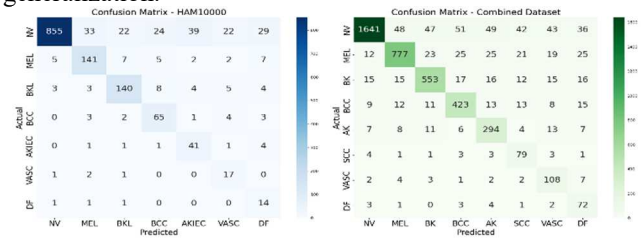

Fig. 4. Confusion matrix of the SkinViT-EfficientX for both datasets.

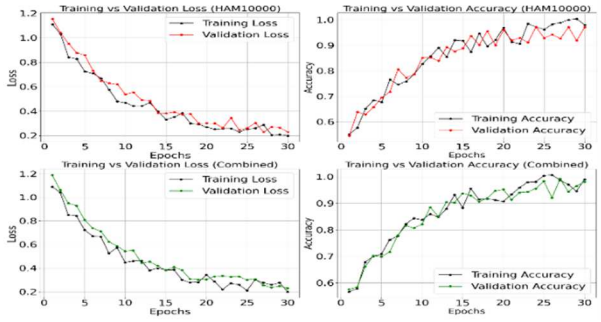

Fig. 5. Learning curve of SkinViT-EfficientX model for both datasets.

C. *State-of-the-art Comparison*

Table II shows a clear evolution in accuracy and practical deployment. VGG16 [19] and DSCCNet [20] achieved 88% and 94.17% accuracy, respectively, but lacked XAI or app integration. XceptionNet [21] improved performance to 90.48%, while the Hybrid CNN [22] reached 93.04% using combined datasets, yet both also missed deployment considerations. CNN+LBP [24] introduced interpretability with 98.41% accuracy but no application. In contrast, the proposed SkinViT-EfficientX models outperformed all prior works (97.93% on HAM10000, 98.86% on Combined) while integrating both XAI and a deployable app, addressing accuracy and usability gaps.

TABLE II. STATE-OF-THE-ART COMPARISON WITH PRIOR STUDIES.

Ref.	Model	Dataset	Result	XAI+App
[19]	VGG16	Skin tumor	88%	No
[20]	DSCCNet	ISIC, HAM10000	94.17%	No
[21]	XceptionNet	HAM10000	90.48%	No
[22]	Hybrid CNN	Combined	93.04	No
[24]	CNN + LBP	HAM10000	98.41%	No
Ours	SkinViT-EfficientX	HAM10000	97.93	Yes
		Combined	98.86	Yes

D. *Explainable Web Application*

Fig. 6 shows a user-friendly interface for a skin lesion classification app. It includes key sections: image upload, Grad-CAM visualization, prediction output, and control actions, improving navigation and efficiency. The "Upload" section allows users to submit dermoscopic images. Next to it, the Grad-CAM output highlights important areas that influence the model's prediction. This feature enhances transparency and trust in the AI's diagnostic reasoning. The heatmap clearly indicates the lesion area critical for classification. The results panel displays the predicted class ("Melanoma") and confidence score (97.6%) in a clear format, enabling quick interpretation for clinicians. A "Clear" button adds usability by allowing users to reset for multiple evaluations easily.

Fig. 6. Explainable web application.

V. DISCUSSION

SkinViT-EfficientX effectively classifies skin lesions in the ISIC 2019 and DermNet datasets, addressing class imbalance and visual similarity better than traditional methods. It uses EfficientNetV2-S for local feature extraction and a lightweight transformer for global dependencies. A confidence-guided token pruning mechanism improves inference efficiency by removing redundant tokens without losing accuracy, making it suitable for real-time clinical use. Preprocessing with the Dull Razor algorithm and anisotropic diffusion filtering removes artifacts and preserves lesion structure, while data augmentation enhances generalization for rare classes like DF, VASC, and SCC. Unlike CNN+LBP and DSCCNet, this model allows for end-to-end learning with Grad-CAM interpretability and features a web-based interface for image uploads, predictions, and heatmap visualizations to support clinical decisions. Limitations include sensitivity to the pruning threshold, underrepresentation of rare classes, and limited validation across various institutions and devices. Future improvements may include adaptive pruning, domain adaptation, federated learning, and mobile deployment to increase accessibility.

VI. CONCLUSION

This study presents SkinViT-EfficientX, a hybrid deep learning framework for multiclass skin cancer classification. Our model uniquely combines convolutional inductive bias with global attention, incorporating a residual cross-attention fusion mechanism and a confidence-guided token pruning strategy to balance accuracy and computational efficiency. We've also added Grad-CAM-based interpretability to increase clinical trust. Evaluated on two benchmark datasets, including ISIC 2019 + DermNet, SkinViT-EfficientX outperformed current state-of-the-art models across several metrics. Its deployment in a real-time web application highlights its potential for teledermatology and point-of-care diagnostics. Our work addresses scalability, interpretability, and multiclass generalization in AI-driven dermatological assessments. Future efforts will focus on cross-device validation, adaptive pruning, and mobile deployment for improved clinical utility and global accessibility.

REFERENCES

[1] B. M. Zugaib Abdalla and C. M. Zugaib Abdalla, "Epidemiology of Skin Cancer," Oncodermatology: An Evidence-Based, Multidisciplinary Approach to Best Practices, pp. 29–35, Jan. 2023, doi: 10.1007/978-3-031-29277-4_3.

[2] R. Archana and P. S. E. Jeevaraj, "Deep learning models for digital image processing: a review," Artif Intell Rev, vol. 57, no. 1, pp. 1–33, Jan. 2024, doi: 10.1007/S10462-023-10631-Z.

[3] Y. Singh et al., "The Role of Geometry in Convolutional Neural Networks for Medical Imaging," Mayo Clinic Proceedings: Digital Health, vol. 1, no. 4, pp. 519–526, Dec. 2023, doi: 10.1016/J.MCPDIG.2023.08.006.

[4] M. M. Rahman, M. Najmul Gony, M. M. Rahman, M. M. Rahman, and S. D. Maria Khatun Shuvra, "Natural language processing in legal document analysis software: A systematic review of current approaches, challenges, and opportunities," International Journal of Innovative Research and Scientific Studies, vol. 8, no. 3, pp. 5026–5042, Jun. 2025, doi: 10.53894/IJIRSS.V8I3.7702.

[5] M. M. Rahman, K. Dhakal, N. Gony, M. K. Shuvra, and M. Rahman, "AI integration in cybersecurity software: Threat detection and response," International Journal of Innovative Research and Scientific Studies, vol. 8, no. 3, pp. 3907–3921, May 2025, doi: 10.53894/IJIRSS.V8I3.7403.

[6] M. Asaduzzaman, K. Dhakal, M. M. Rahman, M. M. Rahman, and S. Nahar, "Optimizing Indoor Positioning in Large Environments: AI," Journal of Information Systems Engineering and Management, vol. 10, no. 48s, pp. 254–260, May 2025, doi: 10.52783/JISEM.V10I48S.9500.

[7] M. Sabbir Hossain et al., "Automatic Navigation and Self-Driving Technology in Agricultural Machinery: A State-of-the-Art Systematic Review," IEEE Access, vol. 13, pp. 94370–94401, 2025, doi: 10.1109/ACCESS.2025.3573324.

[8] M. S. Hossain, M. Rahman, M. Ahmed, M. M. Kabir, M. F. Mridha, and J. Shin, "DeepLabv3Att: Integrating Attention Mechanisms in DeepLabv3 for Enhanced Road Segmentation," 2024 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2024, pp. 711–718, 2024, doi: 10.1109/3ICT64318.2024.10824285.

[9] S. M. S. Hasan, L. Gudala, M. Soni, A. Quraishi, I. Keshta, and F. A. Reegu, "Deep Reinforcement Learning-Based Real-time Data Analytics of Automatic Industrial Robot for High Speed Communication," 2nd IEEE International Conference on Innovations in High-Speed Communication and Signal Processing, IHCSP 2024, 2024, doi: 10.1109/IHCSP63227.2024.10960131.

[10] S. M. Abid, Q. Xiaoping, M. M. Islam, M. A. Islam, M. M. Rahman, and A. R. M. Alam, "Edge-Conditioned GAT for Journal Ranking in Citation Networks," 2025 IEEE 6th International Seminar on Artificial Intelligence, Networking and Information Technology, AINIT 2025, pp. 1612–1619, 2025, doi: 10.1109/AINIT65432.2025.11035687.

[11] A. Pokharel, R. Dantu, S. Zaman, S. Talapuru, and V. Quach, "Deliberation Leads to Unanimous Consensus," 2024 6th International Conference on Blockchain Computing and Applications, BCCA 2024, pp. 356–363, 2024, doi: 10.1109/BCCA62388.2024.10844448.

[12] S. Zaman, R. Dantu, A. Pokharel, S. Talapuru, and V. Quach, "Indivisible State Mirroring in Cross-Chain Atomic Swap across Heterogeneous Blockchains," 2024 6th International Conference on Blockchain Computing and Applications, BCCA 2024, pp. 116–123, 2024, doi: 10.1109/BCCA62388.2024.10844452.

[13] S. Zaman, R. Dantu, S. Badruddoja, S. Talapuru, and K. Upadhyay, "Seamless Asset Exchange in Interconnected Metaverses: Unraveling On-Chain Atomic Swap," Proceedings - 2023 5th IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications, TPS-ISA 2023, pp. 146–155, 2023, doi: 10.1109/TPS-ISA58951.2023.00026.

[14] M. Rahman, M. S. Hossain, A. A. Eva, M. M. Kabir, M. F. Mridha, and J. Shin, "Alzheimers Disease Classification with a Hybrid CNN-SVM Approach on Enhanced MRI Data," 2024 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2024, pp. 50–57, 2024, doi: 10.1109/3ICT64318.2024.10824633.

[15] B. A. Oza, I. R. Khan, N. S. L. K. Kanulla, S. M. S. Hasan, S. K. Davuluri, and L. N. Pothakamuri, "Artificial Intelligence based Ensemble Learning Approach for Brain Tumour Classification Using MRI Images," Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2024, pp. 1583–1588, 2024, doi: 10.1109/IC3I61595.2024.10828861.

[16] M. Rahman, M. S. Hossain, U. Rozario, S. Roy, M. F. Mridha, and N. Dey, "MultiSenseNet: Multi-Modal Deep Learning for Machine Failure Risk Prediction," IEEE Access, vol. 13, pp. 120404–120416, 2025, doi: 10.1109/ACCESS.2025.3586978.

[17] S. M. S. Hasan, I. Keshta, D. Thakkar, L. Gudala, M. Soni, and R. R. Maaliw, "Deep Learning-Based Intrusion Detection Model for Internet of Things Environment," 2nd IEEE International Conference on Innovations in High-Speed Communication and Signal Processing, IHCSP 2024, 2024, doi: 10.1109/IHCSP63227.2024.10960180.

[18] P. Pandey, T. Lawanya, and S. M. Shakib Hasan, "Heat Transfer Analysis of a Stretching Sheet in TRI Particle-Enhanced Nanofluid Systems," FMDB Transactions on Sustainable Applied Sciences, pp. 21–33, Jun. 2024, doi: 10.69888/FTSASS.2024.000164.

[19] S. Bechelli and J. Delhommelle, "Machine Learning and Deep Learning Algorithms for Skin Cancer Classification from Dermoscopic Images," Bioengineering 2022, Vol. 9, Page 97, vol. 9, no. 3, p. 97, Feb. 2022, doi: 10.3390/BIOENGINEERING9030097.

[20] M. Tahir, A. Naeem, H. Malik, J. Tanveer, R. A. Naqvi, and S. W. Lee, "DSCC_Net: Multi-Classification Deep Learning Models for Diagnosing of Skin Cancer Using Dermoscopic Images," Cancers 2023, Vol. 15, Page 2179, vol. 15, no. 7, p. 2179, Apr. 2023, doi: 10.3390/CANCERS15072179.

[21] S. Jain, U. Singhania, B. Tripathy, E. A. Nasr, M. K. Aboudaif, and A. K. Kamrani, "Deep Learning-Based Transfer Learning for Classification of Skin Cancer," Sensors 2021, Vol. 21, Page 8142, vol. 21, no. 23, p. 8142, Dec. 2021, doi: 10.3390/S21238142.

[22] E. Farea, R. A. A. Saleh, H. AbuAlkebash, A. A. R. Farea, and M. A. Al-antari, "A hybrid deep learning skin cancer prediction framework," Engineering Science and Technology, an International Journal, vol. 57, p. 101818, Sep. 2024, doi: 10.1016/J.JESTCH.2024.101818.

[23] A. Panthakkan, S. M. Anzar, S. Jamal, and W. Mansoor, "Concatenated Xception-ResNet50 — A novel hybrid approach for accurate skin cancer prediction," Comput Biol Med, vol. 150, p. 106170, Nov. 2022, doi: 10.1016/J.COMPBIOMED.2022.106170.

[24] F. Grignaffini, E. De Santis, F. Frezza, and A. Rizzi, "An XAI Approach to Melanoma Diagnosis: Explaining the Output of Convolutional Neural Networks with Feature Injection," Information (Switzerland), vol. 15, no. 12, p. 783, Dec. 2024, doi: 10.3390/INFO15120783/S1.

[25] A. A. Abdulredah, M. A. Fadhel, L. Alzubaidi, Y. Duan, M. Kherallah, and F. Charfi, "Towards unbiased skin cancer classification using deep feature fusion," BMC Med Inform Decis Mak, vol. 25, no. 1, pp. 1–22, Dec. 2025, doi: 10.1186/S12911-025-02889-W/FIGURES/16.

[26] A. Halder, A. Dalal, S. Gharami, M. Wozniak, M. F. Ijaz, and P. K. Singh, "A fuzzy rank-based deep ensemble methodology for multi-class skin cancer classification," Sci Rep, vol. 15, no. 1, pp. 1–25, Dec. 2025, doi: 10.1038/S41598-025-90423-3;SUBJMETA=1042,117,258,639,705;KWRD=COMPUTATIONAL+SCIENCE,COMPUTER+SCIENCE,INFORMATION+TECHNOLOGY.

[27] L. Cino, C. Distante, A. Martella, and P. L. Mazzeo, "Skin Lesion Classification Through Test Time Augmentation and Explainable Artificial Intelligence," Journal of Imaging 2025, Vol. 11, Page 15, vol. 11, no. 1, p. 15, Jan. 2025, doi: 10.3390/JIMAGING11010015.

[28] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," Scientific Data 2018 5:1, vol. 5, no. 1, pp. 1–9, Aug. 2018, doi: 10.1038/sdata.2018.161.

[29] E. Emary, H. M. Zawbaa, A. E. Hassanien, G. Schaefer, and A. T. Azar, "Retinal blood vessel segmentation using bee colony optimisation and pattern search," Proceedings of the International Joint Conference on Neural Networks, pp. 1001–1006, Sep. 2014, doi: 10.1109/IJCNN.2014.6889856.

[30] T. A. Rimi, N. Sultana, and M. F. Ahmed Foysal, "Derm-NN: Skin Diseases Detection Using Convolutional Neural Network," Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020, pp. 1205–1209, May 2020, doi: 10.1109/ICICCS48265.2020.9120925.

[31] A. S. U. K. Pranta et al., "A Novel MaxViT Model for Accelerated and Precise Soybean Leaf and Seed Disease Identification," Computers 2025, Vol. 14, Page 197, vol. 14, no. 5, p. 197, May 2025, doi: 10.3390/COMPUTERS14050197.

[32] Md. R. Ahmed et al., "Hierarchical Swin Transformer Ensemble with Explainable AI for Robust and Decentralized Breast Cancer Diagnosis," Bioengineering 2025, Vol. 12, Page 651, vol. 12, no. 6, p. 651, Jun. 2025, doi: 10.3390/BIOENGINEERING12060651.

[33] J. Debnath et al., "LMVT: A hybrid vision transformer with attention mechanisms for efficient and explainable lung cancer diagnosis," Inform Med Unlocked, vol. 57, p. 101669, Jan. 2025, doi: 10.1016/J.IMU.2025.101669.